SCHOOL-BASED EVIDENCE FOR THE VALIDITY OF CURRICULUM-BASED MEASUREMENT NORMS IN SCHOOL DISTRICT #57

by

Saima Fewster

B.G.S., Simon Fraser University, 1987

THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

in

CURRICULUM AND INSTRUCTION

UNIVERSITY OF NORTHERN BRITISH COLUMBIA

March, 2000

ABSTRACT

This study examines the relationship between local Curriculum-Based Measurement (CBM) scores and later academic performance. It adds school-based validity evidence to the body of research literature that has demonstrated CBM to be valid and reliable for a number of assessment purposes.

CBM reading and written expression scores for 678 Grade 6 and 7 students were compared to their year-end English and Social Studies marks received in Grades 8, 9 and 10. Correlational and regression analyses confirmed the predictive validity of CBM scores. Data on the academic programming of the students in secondary school allowed for discriminant analysis procedures that demonstrated the ability of CBM to differentiate between student groups of various proficiency levels. Overall, findings support the appropriateness of the use of CBM norms for local assessment applications, including the use of CBM as an indicator of academic performance when making special education eligibility decisions.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGMENTS

My thanks to Martha Ottesen and Carl Anserello for promoting the awareness and use of Curriculum-Based Measurement in my teaching practices. Thanks to my special education colleagues who constantly encouraged me with their interest in this project, to the school district personnel who participated in the collection of data, and to Lynda and Julie for listening.

I would also like to express my gratitude to Dr. Peter MacMillan for his ongoing assistance, support, and high expectations in the successful completion of this document.

Finally, there is no limit to the amount of appreciation I feel for the patience and consideration provided by my family from the start to the finish of my master's program endeavor. To my husband Rick and my daughters Jennifer and Kristen, I owe my greatest thanks.

CHAPTER ONE: INTRODUCTION

Educators use a variety of assessment tools to collect data that will assist them with the decisions they need to make in their evaluations of student achievement. An important feature of any assessment tool selected is its validity, and developers are responsible for providing evidence, often developed later, to show that the assessment measures what it is intended to measure. Are the results relevant and useful for the purpose intended and how well do they relate to other measures of the same behavior? Educators can choose appropriate assessment tools with greater confidence when this information is provided.

In School District #57, a place has been established for the use of Curriculum-Based Measurement (CBM) assessment procedures, developed and outlined by Stanley Deno (1985). The field of school psychology has acknowledged CBM as an appropriate alternative to traditional assessment methods at the elementary school level (Reschly & Grimes, 1990), and a considerable research base provides evidence that it is a valid and reliable assessment tool. These sources support the implementation of CBM in local special education initiatives.

District personnel first promoted the use of CBM as an effective tool for ongoing measurement of student skills growth. Over time, they also recognized its usefulness for identifying students in need of educational support and its value in

developing effective instructional interventions for those students. Reviews of research identified numerous studies confirming the validity of CBM for these purposes (Good & Jefferson, 1998; Marston, 1989). A discussion of these studies is provided in the literature review chapter of this paper.

As a result of the 1993 School Support Services Task Force recommendations (School District #57, 1994), a formal problem-solving model for the delivery of special education services was adopted in the district. This model followed a number of the principles and processes used by Iowa in a system-wide educational reform initiative implemented in the late 1980s (Tilly & Grimes, 1998). Included in the model was a refined procedure for accessing district-level support resources for students, with CBM incorporated as one criterion for student eligibility. This application of CBM, outlined by Shinn (1989), was facilitated through the development of local CBM norms (School District #57, 1996b). These norms provided a consistent standard for decision-making across schools.

Since the implementation of the district problem-solving model, CBM data has been routinely collected and integrated into school intervention plans. It has been used to give one required measure of the discrepancy between a student's performance and that of same age peers when determining eligibility to apply for district support resources. The intent of the current study is

to add to the information available on the technical adequacy of the CBM procedures adopted, by examining school-based evidence of validity for the CBM scores used to develop local norms. This is consistent with Shinn's (1989) charge that "the importance of quality of the norms is paramount" (p. 114), and with professional guidelines for establishing and reporting the validity of standardized assessment methods.

Research Problem

The local standardization of CBM measures for reading and written expression has generated an assessment tool that relies on technical adequacy data referenced to population samples in the United States. A more acceptable practice when comparing students to standardized norms is to ensure that the norms have been developed from representative samples of the population to which the student belongs. Also, in spite of the numerous well-documented research studies on the validity of CBM procedures, these studies focus primarily on relationships between CBM and standardized achievement measures of reading comprehension or written expression. Less information is available on how CBM measures compare to actual school performance.

Providing evidence to show that the local CBM percentile norms are valid for the purposes that they are being used is important, both to support district use and to increase user

confidence. This study examines evidence of the validity of CBM scores for predicting later junior secondary school achievement, to verify its adequacy as a standardized indicator of student performance. It also examines the appropriateness of using CBM as a basis for the inferences and decisions made about students in a problem-solving model of service delivery, by determining its effectiveness in differentiating between students.

Determining validity in this manner, as a collection of more than one form of supporting evidence, follows assessment guidelines generally adopted within the Canadian educational context (Principles for Fair Student Assessment, 1993). It is also a necessity according to the contemporary perspective provided in Messick's (1989) framework for validation procedures.

Rationale for CBM Validation Study

Whenever comparisons are made between the performance of a student and the student's peers through the use of specific measurement procedures, the technical adequacy of the measure is an issue. This is acknowledged by Deno (1989), who goes on to add that "any time a question arises as to whether or not a performance discrepancy is important, the validity of a particular measurement or set of measurements must be established" (p. 16). Educators who are concerned about the

adequacy of their decision-making practices look for assessment tools that are proven to be both reliable and valid.

Shinn (1989) discusses the appeal of using local norms for procedures such as CBM where "students are assessed using their own curricula and are compared to students receiving the same instruction coming from similar backgrounds and learning experiences" (p. 92). He also stresses the importance of examining and reporting the specific technical features of the norms developed. The adequacy of local norms needs to be carefully examined by both developers and users.

The guidebook containing local norms developed by School District #57 (1996b) reports how CBM validity in general was established, but localized evidence of validity is limited. Although the district has endorsed and encouraged the use of CBM as one assessment tool in the problem-solving process, many teachers remain unconvinced that these seemingly simple little measures can give an accurate indication of a student's level of reading and writing performance. This issue seriously affects the confidence placed in assessment results by teachers and consequently, their willingness to use CBM methods. Building several lines of evidence confirming the validity of CBM scores for the purposes identified may alleviate this problem.

Teachers are required to assign letter grades, based on percentage marks awarded, that indicate a student's achievement

in school subject courses. CBM procedures can be used to evaluate student proficiency in some of the basic skills considered necessary for academic success. If a positive relationship between CBM scores obtained and marks awarded is identified, it can provide school-based confirmation of the CBM validity claims in past research studies. In particular, it would validate CBM scores as indicators, or predictors, of performance in the broader academic domains that include reading and written expression (Good & Jefferson, 1998). Evidence of this relationship would demonstrate a predictive link between the use of CBM scores and actual student achievement outcomes in the form of junior secondary school marks.

According to Salvia and Ysseldyke (1991), screening and eligibility decisions should be based on identifying students who are sufficiently different from their peers to warrant special attention. To further strengthen the claim of validity, evidence to support the use of the CBM norms for this purpose is needed. The ability to differentiate between groups has been reported in CBM literature (Marston, 1989; Shinn, 1989), but validity evidence related to the local population would lend more weight to the use of the locally developed norms.

It is hoped that confirming the validity of CBM will translate into a willingness of more teachers to expand their use of CBM procedures for other purposes. Beyond its application

as a performance assessment tool in the problem solving process, CBM provides current data to assist with setting goals when planning an intervention program (Shinn, Nolet, & Knutson; 1990). It is also a worthwhile method for the ongoing monitoring and evaluation of student progress on a selected intervention plan (Shinn & Bamonto, 1998). In their attempts to develop effective interventions, teachers need a tool that is easy to use, is appropriate to administer frequently, and is sensitive to short term effects. CBM is such a tool, allowing specific evidence of student learning to be detected and documented.

Research Questions and Hypotheses

What is the relative strength of the relationship between oral reading and written expression CBM scores and future general academic outcomes in Humanities courses, as measured by teacher-assigned marks?

Are local CBM procedures valid measures for predicting student achievement in academic courses that most heavily rely on the basic skills of reading and written expression?

To investigate these questions, the following statistical hypothesis has been constructed. It tests conclusions resulting from Pearson's product-moment correlation coefficients between elementary CBM scores and year-end marks in Humanities courses:

H_0: $\rho = 0$

H_1: $\rho \neq 0$

where ρ is the population correlation parameter.

According to conventions for hypothesis testing, the level of significance is set at $\alpha = .05$. Sample sizes for the various correlations range from 212 to 430 students.

Cohen's measure of effect size (1992) identifies the correlation coefficient value that will be considered important. Using Cohen's rule of thumb, the r value of .1 indicates a small-effect size, .3 indicates a medium-effect size and .5 is considered a large-effect size. Following these guidelines, a small effect size of $r = .1$ is not so small as to be considered trivial. A medium effect size represents "an effect likely to be visible to the naked eye of a careful observer" and "approximates the average size of observed effects in various fields (Cohen, p. 156). A large effect size is set the same distance above medium as small is set below it.

Can CBM scores identify a discrepancy in academic achievement between student groups with different levels of performance?

This question is answered by examining the ability to differentiate between identified junior secondary school student programs based on CBM scores, using discriminant analysis procedures. The program groups to be examined include students placed in some form of alternate special education class (Group 1), students receiving remedial support (Group 2), students in

general education classes (Group 3), and students in honours classes (Group 4). Stated statistically, the hypothesis is:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$ where μ_1, μ_2, μ_3, and μ_4 denote the means of the four group populations, and

H_1: $\mu_j \neq \mu_{j'}$ where $j = 1, 2, 3,$ or 4, and $j \neq j'$

Again, the level of significance is set at $\alpha = .05$. The sample size for this analysis is 454 students.

Definition of Terms

Curriculum-Based Measurement (CBM)

Shinn and Bamonto (1998) define CBM as "a set of standard simple, short-duration fluency measures of reading, spelling, written expression, and mathematics computation" (p.1). CBM is a fluency-based model of assessment that is recommended for quantifying student performance changes in the acquisition of basic skills (Choate, Enright, Miller, Poteet & Rakes, 1992). It is also considered as one appropriate method for defining low achievement when making decisions on the referral of students for special education services (Shinn, 1989).

CBM does not, nor was it intended to, sample all the behaviors in an academic domain. Deno (1985) describes CBMs as the indicators of general outcomes, or "the 'vital signs' of student educational health" (p. 230). The importance of the basic skills selected as indicators of learning across many

curriculum areas is well documented (see Choate et al., 1992), and their mastery is a primary focus of school programs.

Outcomes

Outcomes are the results of learning experiences or interactions between students and the educational process.

Indicators

Indicators are numbers or other symbolic representations that can be used to determine whether outcomes have been achieved.

CBM Probes

CBM probes are short, concise curriculum-based measures administered to assess basic skills competence. They primarily consist of the following:

1. In reading, students read passages from their reading curriculum orally. The number of words read correctly and number of errors in a 1-minute interval are counted.
2. In written expression, students write a story for 3 minutes after being given a story starter or topic sentence. The number of words written, words spelled correctly, and correct word sequences are counted.
3. In spelling, students write words that are dictated orally at specified intervals (generally 7 seconds) for 2 minutes. The number of words spelled correctly and correct letter sequences are counted.

4. In math, students write answers to grade-level computation problems during a 2-minute interval. The number of correct digits is counted.

Norms

Norms are converted raw scores, usually appearing in tables, that are established for the purpose of interpreting a particular raw score in terms of its relative location and frequency within the total score distribution (Crocker & Algina, 1986). Normative scores provide information about a student's performance in comparison to a peer reference group. Salvia and Ysseldyke (1991) state that the norms must represent the population on which the test will be used, and they must be current (within 15 years is recommended).

Junior Secondary School

In School District #57, junior secondary school is comprised of the last three years of the British Columbia Intermediate program. This includes Grades 8, 9 and 10. By the end of these grades, students are typically 13, 14, and 15 years of age.

Humanities Courses

In British Columbia, school curriculum is organized according to broad categories, or strands of human experience. The Humanities strand includes the branches of learning that provide "increased opportunity for communication and positive

social interaction, which in turn can lead to global understanding" (British Columbia Ministry of Education, 1990, p. 92). Courses offered within the Humanities strand are English or Language Arts, Social Studies, Learning for Living, and French as a Second Language. For the purpose of this study, Humanities refers to English and Social Studies, the two courses assumed to most heavily rely on the basic skills of reading and writing.

Year End Course Marks

At the end of a secondary school year, teachers assign a letter grade to indicate a student's overall performance in each course taken. Teachers often include the percentage mark awarded to the student. As defined in student reporting guidelines, letter grades are marks corresponding to achieved percentages that "indicate students' levels of performance as they relate to the expected learning outcomes set out in provincial curriculum guides for each subject or course and grade" (British Columbia Ministry of Education, 1994, p. 4). In British Columbia, the percentages associated with letter grades are:

A	86 - 100
B	73 - 85
C+	67 - 72
C	60 - 66
C-	50 - 59
F	0 - 49

These marks are based on teacher judgements of student learning according to specific criteria primarily determined by the teacher using the goals, objectives and expected learning outcomes outlined in the provincial curriculum.

Problem-Solving Model

A problem-solving model provides a systematic framework for making educational decisions about students. In School District #57, a four-level problem-solving model has been implemented for diagnosing student needs and developing intervention plans (School District #57, 1996a). Each successive level in the model suggests a wider range of resources to access for a student (see Appendix A), as resistance to change increases the magnitude of the problem being experienced.

The process at each level in the model follows identical steps, as laid out by Deno (1989) and supported by Salvia and Ysseldyke (1991). These include problem identification and analysis, problem definition or certification, considering alternative solutions in the development of an intervention plan, implementation and monitoring of the plan, and evaluation of the plan to determine when or whether the problem has been solved. At each step, a variety of information needs to be collected to support the decisions made. CBM functions as one procedure adopted for providing objective, reliable, valid and precise data to contribute to informed academic decisions.

At the most formal level of the problem-solving model, the eligibility of a student for additional district resources is made based on documented evidence that the identified problem is pervasive, severe and resistant to intervention attempts made at the previous levels. Again, CBM constitutes one of several acceptable sources of information that can be used for providing this evidence.

Limitations of the Study

This examination of the relationship between CBM scores and subsequent junior secondary school marks does not focus on the many factors that influence student achievement. In addition, it is recognized that the level of basic skills proficiency represented by CBM tasks performed is only one component of all that is required by students to be academically successful. Both of these issues impact on the strength of any relationship evidence found.

The study is limited to data collected for the students that could be located from the CBM norming project. An attrition effect may be present due to the unavailability of information for those students who are no longer in the local school system. In addition, complete information was not always available for the located students due to transfers between schools, variations in school record-keeping systems, and differing data

collection formats utilised by junior secondary school personnel.

The relationship between CBM normative scores and junior secondary school marks is determined by correlation coefficients that are calculated based on the reliability of the two measures. In choosing the school performance measure of course marks as the criterion, the many factors that create possible errors of measurement are acknowledged as limitations to the strength of the predictive validity coefficients calculated. These coefficients may be attenuated, or lower than expected, due to reduced group variability within the student marks. The influence of remedial support already provided to low-achieving students is unknown. Students repeating courses have less data available because higher grade-level courses in those subject areas were not yet taken. Likewise, course marks are not available for students placed in special education programs.

A possible reason for an artificial reduction of group variability in course marks is the problem of varying standards between classes with differing levels of difficulty. Marks for students in transitional or remedial classes may be inflated in comparison to general education student marks, due to less difficult curricular expectations. In a similar manner, marks for students in honours classes may be deflated due to the more advanced content and tasks as compared to courses for general

education students. Another factor that affects variability is the assignation of one overall year-end mark to some students for Grade 8 English and Social Studies courses combined under the heading of Humanities. This is the case in two of the secondary schools providing data for this study.

Finally, this study does not directly address the use of CBM for monitoring student progress or making instructional decisions, even though it is hoped that the results will prompt greater teacher confidence for exploring these applications of CBM procedures. Interested readers are directed to substantial literature available for more information on progress monitoring and program evaluation (see Deno, 1985, 1992; Fuchs, 1989; Shapiro, 1989; Shinn, 1989; Shinn, Nolet, & Knutson, 1990).

CHAPTER TWO: LITERATURE REVIEW

This chapter presents an overview of four major themes. First, information about the history of CBM is summarized to provide the reader with background knowledge on this assessment tool. Next, an explanation of the CBM norming project in School District #57 is presented. Then validity theory is discussed briefly, with a focus on methods for establishing test validity. The fourth section reviews validity evidence for CBM in the literature. The chapter is concluded with an explanation of how the current study contributes to CBM validity research.

Development and Use of Curriculum-Based Measurement

The goal of the work done by Stanley Deno and Phyllis Mirkin in the early 1970s was to provide special education teachers with an efficient, accurate way of assessing the effects of instruction (Shinn & Bamonto, 1998). The result of their efforts was a model that focussed on the frequent administration of short duration skill probes taken from the students' curriculum. These probes provided direct measurement of observable behaviors of interest that could be used as indicators of student progress (Shapiro, 1989).

Deno (1992) describes a 3-step approach that was used for the development of the measurement procedures. First was the identification of a number of alternative behavioral indicators

of the basic skills that might be used in a measurement system. The basic skills identified as measures of student achievement included reading, written expression and spelling. Mathematical computation was later included. Next, measurement formats were created from selected indicators based on the parameters of presentation, duration and source of stimuli. Finally, the potential measures were examined through research and through use with school-aged children.

All research and development activities were guided by a set of design characteristics specified by Deno (1985). He insisted that the measures had to be reliable and valid, simple and efficient, easily understood, and inexpensive. The fluency measures finally chosen as behavioral indicators of these basic skills were selected as a result of the research base initiated at the University of Minnesota. Once selected, Deno again prompted a number of field tests and extensive investigations of the use of these measures. Researchers examined and verified their technical adequacy, creating an extensive research base for CBM implementation.

Salvia and Ysseldyke (1991) describe assessment as a process of collecting data to make decisions about students. They delineate five kinds of assessment decisions including evaluation of the individual, screening, placement, intervention planning, and program evaluation. Along with this overview they

stress that the overall purpose of assessment should be to improve instruction for children. The problem-solving model of service delivery laid out by Deno (1989) is consistent with the procedural format for assessment offered by Salvia and Ysseldyke. Shinn and Hubbard (1992) examined the efficiency and utility of CBM as a common database across all of the decision-making areas of the problem-solving model. Their study reviews research that outlines the flexibility and utility of CBM applications within this context.

The research focusing on specific uses of CBM supports its appropriateness in all stages of problem solving. Marston, Mirkin, and Deno (1984) explain how CBM procedures may reduce assessment bias in problem identification. They also recommend the development of local norms, making differentiation among students possible. This allows for the use of CBM as a viable alternative to traditional norm-referenced tests when screening for eligibility (see also Elliot & Fuchs, 1997; Marston & Magnusson, 1989; Shinn, Nolet, and Knutson, 1990).

Reschly and Grimes (1990) discuss the use of CBM as an alternative assessment method within the field of school psychology. They refer to their 1987 national survey in the United States that indicated two thirds of a school psychologist's time was devoted to various aspects of special education eligibility determination. The clinical procedure of

an intellectual assessment was nearly always one component of that determination. They state that this type of procedure has become increasingly less appropriate in the school setting, adding that "opportunities now exist, as well as the necessary technology, to conduct direct and repeated measures of the behaviors of concern in the natural setting. Practice in the 1990s needs to reflect these changes" (p. 426).

According to Reschly and Grimes (1990), there is a growing dissatisfaction with the dominance of intellectual assessment as standard practice evident in educational psychology literature. This is attributed to the limitations of such assessments in developing educational programs or interventions, improper procedures that introduce test bias and "lost opportunities for the delivery of other, potentially more effective, services due to the excessive time and energy devoted to intellectual assessment" (p. 427). They suggest that the kind of information needed to assist school psychologists with the shift in emphasis from eligibility questions to more functional intervention decisions can be collected through the use of CBM. Shinn, Nolet, and Knutson (1990) agree, stating the CBM embodies preferred educational assessment practices for school psychologists.

Further examinations of CBM have determined this procedure to be helpful in judging placement in programs and setting student goals (Fuchs, 1989). Moreover, studies have shown how

the information collected through CBM procedures can be directly linked to planning interventions and evaluating intervention effectiveness (Fuchs & Fuchs, 1986; Shapiro, 1989; Shinn, Nolet & Knutson, 1990). The opportunity for ongoing monitoring allows teachers to evaluate program effectiveness and make systematic decisions about whether or when to modify a student's instructional program (Allinder, 1996; Fuchs, 1989).

Research in the 1990s has focused on various applications of CBM. Shinn's (1998) text contains chapters built on research related to computer applications, assessing early literacy skills, reintegration of students into general education programs, and use of CBM with minority and secondary students. Going beyond the original basal reading material focus for CBM, reading performance in content area texts (Espin & Foegen, 1996) has been studied, as has its use with authentic reading material (Hintze, Shapiro, Conte, & Basile, 1997). In School District #57, Hedekar (1997) examined the effects of age and gender on CBM scores. Other current topics discussed in literature on CBM use include teacher efficacy and the improvement of instruction, increasing student learning and motivation, and program evaluation.

Shinn and Bamonto (1998) provide an analysis of the few criticisms of CBM published. These criticisms are attributed primarily to a lack of understanding about CBM technical issues,

philosophical differences of opinion related to assessment, and "the inevitable resistance to change in assessment practices and corresponding changes in conceptions of what defines a problem" (p. 27). Some professionals with relatively little information about CBM understandably find themselves questioning an unfamiliar procedure due to working from a limited knowledge base. Model confusion between CBM and CBA (Curriculum Based Assessment) add to the questions raised.

Of note are the concerns discussed which echo the criticisms of CBM voiced by general educators. One concern is a lack of confidence that the seemingly simple fluency probes provide a meaningful measure of the intended behavior. Another is the quantitative and prescriptive nature of CBM. Also prevalent are the erroneous beliefs that CBM will be accepted as the only measure of student achievement or that it is intended to be used to assess every academic content domain. In their endeavor to improve the understanding of CBM, Shinn and Bamonto (1998) address these and other issues by responding with detailed explanations based on research. Ultimately, their goal in addressing CBM issues is to facilitate changes to assessment practices that will mean an improvement in the quality of services to students.

General acceptance and adoption of CBM has spread from Deno's (1985) original efforts in Minnesota to many school

districts throughout the United States (Shinn and Bamonto, 1998). Both general education and special education classrooms effectively apply CBM procedures as common practice. The US Department of Education has supported its use, providing federal funding for a number of related research and development projects. Tilly and Grimes (1998) discuss the statewide endorsement and use of CBM in Iowa's reform of its special education delivery system. In the years since its inception, the assessment procedures known as Curriculum-Based Measurement have been taken continuously through the process of development, implementation, evaluation and revision. Research studies have repeatedly focused on the reliability, validity and effective use of CBM.

According to Holmes (1993), the improvement of learning as a result of educational measurement needs to meet two conditions. First, the outcomes being tested must be recognized and accepted as important objectives of the instructional program. CBM, as an indicator of general outcomes, focuses on basic skills commonly identified as important for academic achievement. Second, achievement assessment must be planned and implemented as an integral part of the curriculum and program of instruction. Again, CBM has been shown to fulfill this requirement in its ability to provide data that can be used in all stages of a problem-solving model.

Currently, CBM applications are being routinely used by a wide range of individual special education and general education teachers, as well as school psychologists interested in alternative approaches to traditional assessment practices. In School District #57, an ideal match was seen between CBM and the principles, guidelines and recommendations that drove the district's adoption of a collaborative problem-solving model for providing support services (School District #57, 1996a). CBM met the desire of district personnel to link functional assessments with effective interventions.

District CBM Norming Project

The 1993 School Support Services Task Force recommendations (School District #57, 1994) set foundation principles that provided for the organization and delivery of school support services (School District #57, 1996a). A "collaborative problem-solving model, which links functional assessments to effective interventions" (School District #57, 1996b, p.3) was mandated. Consistent with this principle, reviews of research literature by district personnel prompted the adoption of CBM as an integral assessment component of the reform initiative for support services delivery.

In order to facilitate the use of CBM by teachers, School District #57 developed local norms in the 1995-1996 school year

for CBM scores in reading and writing. The intent of the norming project was to provide a locally standardized tool for reading and writing measures to support the district's model of service delivery. This assessment tool allowed for student performance comparisons as one measure of the severity of a learning problem when determining a student's eligibility for district support resources.

Personnel in each elementary school in the district were trained in the administration and scoring of CBM reading and writing probes. A project manual provided background information to school staff and outlined all procedures to ensure that standardization issues were understood (School District #57, 1995). Schools collected CBM scores over three assessment periods, providing the database for the norming project.

Once collected, all CBM test scores were given to the University of Northern British Columbia for processing under the direction of Dr. Peter MacMillan. The resulting technical information was compiled in a guidebook for the use of the CBM norms (School District #57, 1996b). The guidebook and district expectations for the implementation and interpretation of CBM procedures were outlined in meetings with administrators and support personnel from each elementary school.

The main features of the guidebook include an introduction to CBM, an explanation of the use of CBM to determine

eligibility for additional district resources, norms tables, and a technical summary report. The section on CBM use provides information instructing the reader on how to apply CBM procedures. The norms tables give percentile ranks for reading and written expression scores, arranged by grade level from Grade 1 to Grade 7, along with instructions for their use. These tables also indicate cut-off scores that place student scores within ranges identified as well above average, above average, average, below average, and well below average.

The summary report briefly presents findings in the literature review done in the district prior to the norming project. It discusses CBM reliability and validity information from previous research studies. The technical adequacy of the data from the norming project is addressed. Descriptive statistics are presented for reading and written expression scores at each grade level. Reliability is reported, with the inclusion of coefficients demonstrating stability over time and across scorers. In addition, validity coefficients are provided to demonstrate that an expected relationship between the reading and writing tasks was found.

Validity Theory

Validation is "the process by which a test developer or test user collects evidence to support the types of inferences

that are to be drawn from test scores" (Crocker & Algina, 1986, p. 217). It is important to know what the test measures and how well that measurement task is accomplished. Is the test meaningful, useful or relevant for making decisions related to a particular purpose? This question is answered by examining the relationships between test performance and other independently observable facts about the behavior characteristics under consideration (Anastasi, 1988). Because validity is such an important characteristic of a test, it is critical that both the developers and users carefully consider all the validation evidence available to them.

Importance of Reliability to Validity

Although validity is considered an essential test quality, it cannot be achieved without the necessary precondition of reliability (Thorndike, 1997). Tests can be very reliable without being valid, but they cannot be valid without being reliable. This means that it is possible for a test to measure some trait very consistently, or reliably, even if the resulting information is not considered important. However, the test cannot be judged as accurate, or valid, if results do not provide consistently similar information. Because tests and other measurement procedures are used to help make decisions, it is desirable for them to be both as accurate and as precise as

possible. Reliability, then, contributes valuable information to the validation process.

"Reliability describes the extent to which measurements can be depended on to provide consistent, unambiguous information" (Sax, 1997, p.271). Although a person's true score is constant, each observation or measurement of that score can be different. Any differences between scores are inconsistencies that result in errors of measurement. Errors can be the result of numerous factors including changes in the person, characteristics of the test itself, or variations in administration or scoring procedures. The reliability of test scores will be high if the scores are repeatable and consistent, and errors of measurement are held to a minimum.

Reliability can be estimated by the procedures of stability, equivalence, stability and equivalence and internal consistency (Sax, 1997). A measure or test is repeated to assess stability. Equivalence is estimated by administering an equivalent test form concurrently. Stability and equivalence are determined by administering two or more equivalent forms at different times. Internal consistency is determined by subdividing one administration of a test into two or more equivalent parts for comparison. The most rigorous reliability standard is met by evidence of stability and equivalence, where

consistent results on parallel forms separated by a time interval are obtained.

Reliability can be expressed in two ways (Thorndike, 1997). One is to describe the amount of variation, or the standard error of measurement, that might be expected in repeated measures of an individual. The other is to determine the correlation coefficient for the degree that the ranking of an individual will be consistent between measures. The standard error of measurement, along with its confidence band, is most appropriate when an index of consistency for individual scores is wanted. The reliability coefficient is more useful for making comparisons between measures, particularly when scores are expressed in different units.

An acceptable level of reliability for a test or measure is determined by a number of factors, and no single interpretation is applicable for all situations. The purpose of the test and the intended use of the scores need to be considered. This may include a decision on whether current or future levels of performance are being assessed. Another issue is whether the test will be used to make decisions about groups or individuals. Even issues of practicality such as test cost, the length of the test, and convenience of administration have a bearing on the level of reliability accepted.

The maximum validity coefficient possible is calculated from the reliability coefficient. Because validity coefficients are directly related to reliability values, they will be lower or higher to the extent that the measures being correlated are estimated as reliable. The size of the estimated reliability coefficient will depend on the potential sources of error affecting the scores obtained. No measurement is perfect and no score is error free. A reasonable way of dealing with the errors of measurement that attenuate, or lower, correlation coefficients is to perform a correction for attenuation. This correction reduces the influence of random errors that may cause a measure to appear invalid, when in fact it is the reliability that is in question. Alternately, correcting for attenuation has little effect on validity when the measure is already reliable.

Reliability of Study Measures

The reliability of CBM scores used in this study was derived from the technical data reported for the School District #57 norming project (School District #57, 1996b). Interscorer reliability was found to range from coefficients of .97 to .99 for reading probes, and an interscorer reliability coefficient of .98 was found for written expression probes. Tables of additional reliability information from the school district's CBM guidebook are included in Appendix B. This information demonstrates the degree of equivalence between measures and

stability over time found for the CBM measures used, and indicates that the measures behaved as expected.

In the early 1900s, junior secondary school teachers introduced percentages as a way to indicate the accomplishments of an increasing number of students in specific subject areas (Guskey, 1994). Guskey reports that these percentage marks have been challenged as reliable measures of student achievement. A criticism of these grading practices is that they are too simplistic and they subject all students to universal standards. Although they are intended to indicate the level of achievement attained, a wide range of information and all the complexity of student learning is reduced into a single grade or percentage. The result is one symbol that reflects great inconsistency due to a wide range of factors. Subjectivity on the part of teachers, interrater agreement, student variability, and variation in the conditions under which marks are provided both between and within schools all contribute to uncertainty about the reliability of marks awarded. An examination of the reliability of marks used in this study was conducted and is presented in later chapters.

Traditional Validation Methods

Studies of content-related validity, criterion-related validity (concurrent and predictive), and construct validity are the most common methods traditionally used to support the types

of inferences to be made from test scores. Other types of validity evidence that can also contribute to the acceptance of an assessment tool are similar to content validity. These include face validity, social validity, curricular validity, and instructional validity.

Judged Validity Evidence

Content-related validity refers to the extent that the objectives or content of instruction are represented on a test, or how accurately the test measures what was taught. Independent experts carefully examine test items and judge their adequacy as samples of the particular domain being assessed. This process includes decisions such as whether or not to weight certain objectives, what aspects of the items to examine, how to structure the student task, student characteristics that may affect the meaningfulness of the judgements made, and how the data gathered should be summarized (Crocker & Algina, 1986).

Determination of the content validity of an assessment tool is judgmental in nature. It requires a clear definition of the content being tested and relies on the expertise and agreement of the judges chosen. The use of the test is supported if the judges deem that the evidence presented accurately reflects the required relevance and representativeness of the content.

Face validity refers to the extent to which a test appears to measure something meaningful to the untrained eye. Although

face validity is not really validity in a technical sense, the cooperation and motivation of examinees can be seriously affected if it is their perception that the test is not a worthwhile measure. As well, user acceptance of the results can be affected if the test is not judged to be relevant to its objectives. Even though face validity is of limited technical importance, it is the type often sought as a substitute for content validity by teachers and administrators (Holmes, 1993).

Three other validity concepts mentioned operate similarly to face validity. Social validity relates to the extent to which the test will be used, based on the perception of the meaningfulness or usefulness of the information provided (Salvia & Ysseldyke, 1991). Curricular validity is used to describe the relevance of test items to the formally described objectives of a specific institution's curriculum, while instructional validity refers to the extent to which teachers have provided instruction in the specific content and skills measured by the items on the test (Crocker & Algina, 1986). These concepts all focus on relevance as an important consideration in determining test validity.

Empirical Validity Evidence

A test has criterion-related validity if the test scores are shown to be related to a behavior on some performance criterion other than the test itself. One of the most difficult

tasks in establishing criterion-related validity "is that of locating or creating a satisfactory measure of success to be used as a criterion measure for test validation" (Thorndike, 1997, p. 144). In many instances there are no objective records of performance, or performance is influenced by several factors that can't be controlled. Often the best that an examiner can do is to choose the most suitable criterion measure from those available. Thorndike identifies the desired qualities of the criterion measure to be relevance, freedom from bias, reliability, and availability.

The two types of criterion-related validity are concurrent validity and predictive validity. Both determine the extent to which a person's score on a criterion measure can be estimated from that person's test score. Concurrent validity is the relationship between test scores and criterion measurements made at the time of the test. Predictive validity is the degree to which test scores predict criterion performance that will be measured at some time in the future.

A validity coefficient describes the relationship between test scores and criterion measures. The importance of the correlation coefficient is determined not only by its magnitude, its statistical significance and the margin of error expected, but also the purpose for which it will be used. Validity coefficients for educational and psychological tests often seem

low, but tests with values of .40 and higher can contribute useful information for making decisions about students (McDaniel, 1994). The predictive validity of most tests is less than .60. According to Anastasi (1988), any significant correlation with the criterion, however low, can indicate that a test appreciably improves predictive efficiency, and even coefficient values as low as .20 or .30 can justify using tests for selection purposes if they are statistically significant at an acceptable level.

Constructs can be described as the general traits or qualities of an individual that are being assessed. They are not observable, but are derived from observations of related variables. Construct validation requires an examination of how accurately the test measures the underlying theoretical assumptions. Indirect evidence is compiled and studied in order to demonstrate that a test acts in the way that is expected and measures the construct that it claims to measure. This information is used to answer questions about what kinds of inferences are supported by obtained test scores.

Current Validity Perspectives

A more contemporary conceptualization of validity suggests that multiple types of evidence are required by all tests. These types of evidence are not alternatives, but are supplementary to each other. Further, validity must be established for each

specific use of a test. Shepard (1993) echoes many other current writers in the measurement field, stating that every test use involves inferences or interpretations that need to be supported by a combination of logical argument and empirical evidence.

Validity is described by Messick (1989) as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (p. 13). This definition stresses that conventional validity measures are desirable, but not sufficient for determining the validity of a measure. Also necessary to consider are the relevance and utility of the measure to provide an adequate basis for decision-making, the values implications that outline whether it is appropriate to make the inference the test implies, and the social consequences of taking action or making decisions based on the test results.

Messick (1989, 1995) maintains that validity is a unitary concept. It encompasses different types of validation procedures that support one unified notion of validity under the heading of construct validity. Within this framework, seeking evidence that supports proposed score inferences, interpretations, or test use (convergent evidence) is only a component of the validation focus. Along with this supporting information, evidence must

also be collected to establish that alternative inferences are not as well supported (discriminant evidence).

Validation is considered an ongoing process by Messick (1995), with existing construct validity evidence becoming enhanced or contravened by new findings. Because validity evidence is always incomplete, validation involves making the most reasonable case possible, based on the evidence available. Almost any kind of information about a test can contribute to and enhance an understanding of score meaning, but the contribution becomes stronger when it is examined for fit with the underlying theoretical rationale for score interpretation.

Along with the traditional forms of validity subsumed under the heading of construct validity, Messick (1995) stresses the importance of examining the social consequences of the test interpretation and use. Potential as well as actual consequences are necessary to consider, in order to capitalize on positive effects, or forestall negative effects, related to the testing. The consequences must be not only supportive of the intended testing purposes but also consistent with other social values. Again, this form of evidence is not viewed in isolation by Messick, but is considered as yet another aspect of construct validity. Consequences and values are viewed as facets of validity. Validity, then, is an integrative summary of evidence and argument supporting test use for the purpose intended.

In response to the complexity of the validity framework presented by Messick (1989), the issue of how much evidence can be considered enough is raised by Shepard (1993). Although she agrees that validity must be established for each particular use of a test, she suggests that researchers set priorities. The focus should be to address the most relevant validity issues related to the test use by asking the question: "What does the test claim to do?" (Shepard, 1993, p. 429). Evidence is then collected to support or refute this claim.

This expanded view of validity is consistent with current assessment guidelines in Canada for test developers to "provide evidence that the assessment method yields results that satisfy its intended purpose(s)" (Principles for Fair Student Assessment, 1993, p.16). These guidelines indicate that assessment methods should be:

1. Clearly linked to the purposes for which inferences and decisions are to be made.
2. Clearly related to the goals and objectives of instruction.
3. Chosen with consideration given to the consequences of the decisions to be made.
4. Supported through the use of more than one assessment method to provide more comprehensive, reliable results.

5. Free from bias and suited to the backgrounds and prior experiences of students.
6. Free from discriminatory content or language.

Meeting the requirements laid out in these principles builds a solid validity argument that encompasses both traditional and current views of validity.

Curriculum-Based Measurement Validity Research

Since the inception of CBM, its technical adequacy has been studied extensively. The many studies of the psychometric properties of CBM have provided a substantial research base that is one of the key advantages of this assessment tool. In addition to confirming its usefulness for assessing basic skills instruction, the versatility and validity of CBM has been investigated for a wide range of other assessment applications.

Marston (1989) provided an exhaustive review detailing research evidence supporting the concurrent criterion-related validity of CBM. He reported the validity and reliability of CBM procedures by summarizing all relevant available studies carried out during the 1980s. In general, information most readily available to Marston was about CBM reading, although written expression studies and, to a lesser extent, spelling and math studies were also discussed. His review includes the technical data and results reported for the various studies.

In his discussion of study results, Marston (1989) indicated several important findings. First, criterion-related validity was demonstrated between reading fluency measures and both criterion-referenced mastery tests of different basal reading series and standardized tests of reading competency (Deno, Mirkin, & Chiang, 1982). In addition, reading fluency measures were found to be highly related to teacher's judgment of student reading proficiency (Deno, 1985; Shinn, Ysseldyke, Deno, & Tindal, 1986).

Marston (1989) also found that oral reading fluency was determined to be a valid measure of reading comprehension in studies examining the relationships between various measures of comprehension (Fuchs, Fuchs, & Maxwell, 1988). Finally, he discussed discriminant analysis studies providing construct validity evidence that CBM procedures performed as well as traditional measures of aptitude-achievement discrepancy for predicting a learning disability classification (Shinn & Marston, 1985).

In particular, studies of the relationship between CBM and other reading measures reviewed by Marston (1989) demonstrated the criterion-related validity of oral reading fluency as an adequate measure of reading. Studies reported validity coefficients in the .60 to .90 range, with most coefficients being above .80. Examinations of reliability for CBM reading

measures reported coefficients ranging from .82 to .97 for test-retest estimates, .84 to .96 for parallel form estimates, and interrater agreement coefficients of .99.

The written expression studies reviewed by Marston (1989) considered comparisons of CBM with criterion measures including teacher holistic ratings of writing skill and published writing achievement tests. These studies reported validity coefficients in the .60 to .80 range (see Deno, Marston, & Mirkin, 1982). Reliability estimates using test-retest and parallel form methods ranged from .41 to .96. A mean of .98 was found for interrater scoring agreement. One study by Fuchs, Deno and Marston (1983) indicated that reliability could be greatly improved by aggregating over samples and forms, resulting in an increase of stability coefficients from .55 for two samples to .89 for 10 samples. As with CBM reading measures, it was established that the differentiation of academic performance between groups of students could be determined with CBM written expression procedures.

In his discussion about direct assessment of academic skills, Shapiro (1989) recommended the use of CBM, referring to the "large, extensive, and impressive data base that substantiates the value of this system" (p. 16). He referred to numerous studies that justify the use of CBM in the evaluation of academic problems. His focus was on evidence validating CBM

as one assessment technique that is useful in all aspects of problem-solving models for service delivery. Deno (1989) expanded on this information, outlining in more detail how CBM fits with the problem-solving model functions of screening and eligibility decision-making, progress monitoring, and program evaluation. Fuchs and Fuchs (1986) discussed how the use of CBM in instructional programs could result in greater student achievement, improved teacher decision-making, and enhanced student awareness of their own performance.

The use of CBM within a problem-solving model of service delivery has received considerable attention. A wealth of research has been published to support the use of CBM beyond the evaluation of effectiveness of instructional interventions for individual students. Shinn & Hubbard (1992) presented a detailed discussion of CBM use in all steps of the problem-solving process, with considerable emphasis on both content validity and the development of a cohesive rationale for use at each level. In their research study, Marston and Magnusson (1985) demonstrated criterion-related validity for CBM through correlations with alternate achievement measures and teacher judgment. They also provided an analysis that indicated significant results, $F(2,269) = 111.8$, $p < .001$, for the differentiation of students by program placement and by discrepancy from peers.

In particular, the validity of CBM use for screening and determining special program eligibility as an alternative to traditional intelligence and achievement tests has been explored. In research by Marston, Mirkin, and Deno (1984), CBM compared favorably with traditional teacher-referral procedures. The study also demonstrated the effectiveness of CBM for negating the influence of biasing factors.

Schendel and Binder-Reschly (1989) added to the criterion-related evidence supporting CBM for this purpose using correlational, discrepancy, discriminant function, and regression analyses. They concluded that CBM is as effective as other traditional psychometric measures, with the added advantage of being readily available, time and cost efficient, and closely tied to program planning and implementation. More recent studies reaffirm these earlier results (Elliot & Fuchs, 1997; Fuchs & Fuchs, 1990; Shinn & Habedank, 1992).

The validity of oral reading fluency measures has continued to receive much attention. Despite the considerable validation information published, educators continue to question the adequacy of CBM for measuring reading skill. In their analysis of this controversy, Shinn, Good, Knutson, Tilly, and Collins (1992) summarized research in the professional literature that points to the importance of fluent, automatic decoding as a fundamental skill for reading comprehension. Competent readers

who allocate less attention to the decoding process are able to devote more attention to processing meaning. Research also indicated that when students' reading rates improved, their comprehension improved (Shinn, Ysseldyke, Deno, & Tindal, 1986). The efficacy of oral reading fluency measures as indicators of reading proficiency and comprehension was strongly supported.

Shinn et al. (1992) used confirmatory factor analysis procedures to demonstrate construct validity for CBM reading fluency measures. Regardless of which factor model was employed in their study, oral reading fluency measures provided a good index of reading proficiency, including comprehension. They suggested that this demonstration of a valid tie between oral reading fluency and theoretical reading process models should put face validity arguments regarding this CBM measure to rest.

The data collected by Espin and Deno (1993) supported the validity of reading aloud in predicting the academic success of students at the secondary level. Their study reported low-moderate to moderately high correlations between CBM and concurrent measures including reading subtests on a standardized achievement test, classroom-based academic performance measures, and grade point averages. Results highlighted the contribution of general reading skill to academic success. The measures were found to be particularly useful for identifying low-achieving students who were likely to experience difficulties in the

content curriculum areas. Espin and Deno also suggested that reading proficiency may become less important as students become better readers and other factors such as motivation, study skills, and background knowledge come into play.

Discussing the validity of oral reading measures with authentic reading materials, Hintze et al. (1997) extended the work of Fuchs and Deno (1992). As in the earlier study, they found that "the sensitivity of the oral reading metric was equally robust regardless of the type of basal (i.e., literature-based or traditional skills-based) used for instruction" (p. 537). Correlation values with large effect sizes (R^2 = .49 and .52) indicated that student reading skill in both types of materials were strong predictors of reading comprehension skills. This information supports research confirming the validity of oral reading fluency as a measure of reading competence and suggests that evaluating comprehension may be redundant to assessing oral reading fluency.

Good and Jefferson (1998) provided another summary of CBM validity research studies related to reading, math, and written expression measures. Their work went beyond examinations of criterion-related validity for CBM to examinations of its construct validity. Information from Marston's review (1989) of concurrent, criterion-related validity studies was combined with other various, more current, research findings. They explained

that although establishing correlations of a variety of measures of the same underlying construct is one way to strengthen the construct validity of a test, it is not enough. Their investigation provided a persuasive argument for the construct validity of CBM. Although construct validity was supported for all measures, less evidence was available for math and written expression, and the magnitude of validity correlations was not as high as for reading.

The approach taken by Good and Jefferson (1998) was to argue that if CBM reading and another test of reading achievement are highly correlated, then both are measuring something in common. That "something" is the construct of interest, and since the construct itself is not directly measurable, correlations provide necessary indirect evidence. The correlation alone is not as important as the evidence that both measures are related to the same underlying construct. Their findings from correlations of a variety of measures indicated that "CBM reading probes are as valid or more valid indicators of reading competence as other available reading measures" (p. 67).

Beyond their discussion of construct validity, Good and Jefferson (1998) also examined the relevance, utility, values implications and social consequences of CBM use in a problem-solving model. Rational arguments presented follow the

validation requirements set out in Messick's (1989) framework, addressing the need to collect a variety of evidence when judging the appropriateness of an assessment tool for the purpose intended. Gersten, Keating and Irvin (1995) acknowledged similar issues in their examination of CBM procedures, as did Knutson & Shinn (1991). Each of these studies confirmed the appropriateness of the use of CBM in the various stages of problem-solving models.

In summary, research has demonstrated the technical adequacy of CBM, providing validity evidence that meets traditional guidelines as well as Messick's (1989) criteria for a more comprehensive validity judgement of assessment appropriateness, meaningfulness and usefulness.

Contribution of This Study to the Literature

The current study is primarily an empirical, predictive validity study, examining the relationship between elementary CBM scores and year-end junior secondary school marks. It follows procedures set out for criterion-related predictive validity in traditional validity theory. It takes into consideration the various guidelines provided for research establishing the validity of locally developed assessment tools.

The evidence gathered also addresses additional requirements from more current conceptualizations of validity,

including the positions taken by Messick (1989) and the Joint Advisory Committee for professional organizations in the Canadian educational context (Principles for Fair Student Assessment, 1993). These positions focus on examining the adequacy and appropriateness of both the assessment methods and the decisions made based on the assessment results. The validity information provided here is a collection of multiple lines of evidence showing that CBM is not only technically adequate, but also appropriate for the intended local purpose.

Assessment tools should provide as much technical data as possible to potential test users. Verifying the validity of the local CBM reading and written expression norms for the purpose of determining significant discrepancies through student performance comparisons adds a missing component to the technical information currently provided along with the norms tables. It also adds to the body of research focusing on assessment practices for the purpose of differentiating students with achievement problems.

Studies of CBM criterion-related validity have primarily focussed on students at the elementary school level, comparing their CBM scores to scores on recognized standardized assessment tools. Results of the data analyses in this study offer CBM validation evidence that has little set precedent in previous literature. The relationship of CBM with school-based measures

of academic achievement has received little attention beyond the study by Espin and Deno (1993) that provided an examination of the relationship between reading and indicators of academic achievement for secondary students. In addition, examinations of CBM validity over time, such as are used in this study, are not readily available, if at all. The predictive ability of CBM scores in School District #57 is determined in this study as a result of school-based evidence from longitudinal data, offering new information to CBM validity research.

A point raised by Shinn and Bamonto (1998) is that professional writing and research on CBM continues to be produced primarily by a few of the original CBM authors. They note that although over 150 articles on CBM have been published since 1988, there has been a lack of independent research. This study offers further validation information to add to the general body of knowledge related to CBM from an independent source.

CHAPTER THREE: METHOD

Research Design

The current study was designed to examine the strength of the relationship between students' CBM scores in the elementary school CBM norming project and their later junior secondary school academic achievement. Predictor variables were the end-of-year CBM reading and written expression scores of students who were in Grade 6 or Grade 7 at the time of the norming project. Year-end percentage marks received by the identified students in their Grade 8, Grade 9 and Grade 10 Humanities courses (English and Social Studies) were anticipated to be, and selected as, relevant criterion measures of academic achievement. This decision reflected a belief that achievement in these courses is often dependent upon student proficiency in the basic skills of reading and written expression. CBM scores and junior secondary school end-of-year percentage marks were analyzed for significant relationships.

Additional indicators of academic performance at the junior secondary school level included the provision of remedial support and student placement in alternate special education classes or honours classes. Most schools included this information along with student marks. These variables offered an opportunity to examine the validity of CBM scores for the purpose of differentiation between groups of students by

academic performance, to confirm its use as a screening and special education eligibility assessment tool. The information was used to categorize students into four groups. Group 1 included students placed in some form of alternate special education class, Group 2 included students receiving remedial support, Group 3 included students in general education classes, and Group 4 included students in honours classes.

Subjects

Student files for the present study were selected from the data set generated by the CBM norming project in the 1995-1996 school year. For the original data collection, approximately 20% of the district students from Grade 1 to Grade 7 were selected through stratified random procedures, with proportional representation from each elementary school. Excluded from the study were Level 1 and Level 2 ESL students, students with intellectual disabilities, and other "hard labeled" students (hearing impaired, visually impaired, autistic, multiple disabilities). Students with learning disabilities were included.

Reading and written expression probes were administered to the students three times during the year. In order to maintain the original sample size, selected students who left the schools were replaced, again following the specified selection procedures (School District #57, 1995). For each student, the

database included name, age, birth date and gender, in addition to the specific CBM probes administered and scores received in each norming period.

A total of 639 students in the district CBM norming sample were taken from Grades 6 and 7. These students were selected for the present study. At the end of the 1998-1999 school year, records for 465 of the students were located in district junior secondary schools. This study is based on the information collected from the files of these students. No information is available for the missing students. It is not known whether they left school, left the district, or were simply not located due to transferring between junior secondary schools. In addition, 25 students identified by first name and last initial only in the original project could not be traced in junior secondary schools and could not be included in this study.

Ethics

A preliminary overview of the research proposal was presented to Norm Monroe, the Director of School Services for School District #57 (see Appendix C). With his approval, introductory letters were sent to all junior secondary school principals to inform them of the project and allow them the opportunity to provide input or obtain clarification as needed (see Appendix D). The study was considered to be a district research project, as it was to be carried out within the

guidelines of normal district practices, with the results of interest to district administration.

The UNBC Ethics Committee approved the research proposal for this study. Dr. Peter MacMillan, as the thesis research supervisor, and thesis committee members, Dr. Bryan Hartman and Dr. Tom Strong, also reviewed and approved the plan.

Dr. Peter MacMillan at UNBC maintains a secured database identifying students from the CBM project. With verification of school district and UNBC approval to proceed, he gave consent to access the names of the Grade 6 and Grade 7 students for the purpose of this study, and a data file of CBM scores for these students was created. Once junior secondary school marks data were collected, student names were removed from all documentation for further phases of the research study. To ensure confidentiality, coded computer data is kept in secured files. The school district has no access to any information related to individual students in the study.

As data collection involved examining existing school records only, there was no direct contact with the students and they were not personally affected by the study in any way. There was no involvement of the teachers who awarded marks to students. Teachers did not know the CBM scores obtained by their students in the elementary norming project. These conditions ensured that there would be no system effect in data collection.

According to school district procedures, the examination of student records for administrative purposes is normal practice and parental consent was not required. Opportunities were provided for individual school administrators to discuss any concerns related to ethics. Each administrator had the right to deny access to student files for their school.

Materials

CBM Measures

CBM probes used in the original district CBM norming project were developed from local curriculum materials according to procedures outlined by Shinn (1989). All students in the project were administered a reading fluency probe and a written expression probe in each of three testing periods during the 1995-1996 school year (October, January and April). Teachers responsible for testing within schools were given a training session on administration and scoring procedures. Reading probes were scored for Words Read Correctly (WRC) and written expression probes were scored for Total Words Written (TWW) and Words Spelled Correctly (WSC).

Reliability of the measures over time and across testers was established and reported (School District #57, 1996b). This information also included evidence of convergent and discriminant validity, showing that the measures behaved as

expected when correlated with each other and when correlated across norming periods.

Junior Secondary School Marks

Student marks for junior secondary school courses were generated by schools in a number of different formats. School administrators decided how the information would be collected in their schools, resulting in computer-generated mark lists, photocopies of student permanent records or edited file summaries. Information may have been lost due to these variations, as some data formats provided incomplete data across all grades for a few students.

Marks were generally provided as year-end percentages, although some were reported as letter grades. When information was presented in the letter grade format, it was converted to a percentage. This was done by selecting the median scores in the percentage ranges set for the letter grades A, B, C+, C, and C-, as determined by the B. C. Ministry of Education (1994). For letter grades reported as P (Pass), the percentage mark was set at the pass-fail cutoff score of 50. Letter grades reported as F (Fail) were given a percentage of 39, the median score of all reported failing percentage scores. With the setting of these marks, there was an assumption that individual over-estimations or under-estimations should offset each other. However, it was

also recognized that some mark variability would be lost and could affect analysis results.

Procedures

Data Collection

Junior secondary school administrators generated lists of the students registered in Grades 9 and 10 at their school. The Grade 6 and Grade 7 students from the CBM project selected for this study were located on the lists provided. Names of the identified students registered at their schools were returned to administrators, who arranged for student marks to be collected and submitted. Accumulated junior secondary school marks for identified students were collected from permanent record files, maintained in computer or hard copy databases in junior secondary school offices. Designated personnel at the schools collected the information from these files. Information for students who transferred between schools was limited, and no marks were available for students gone from the district or no longer attending school.

Student marks in English and Social Studies courses were entered as percentage scores into the computer data file that included CBM scores from the norming project. Marks for repeated courses were not included. A group variable was created to indicate the type of junior secondary school programming

provided to the students during their two or three years of attendance. This information was categorized into four groups. Group 1 included students who had been placed in some form of an alternate special education program. Group 2 included students who received remedial support, either in the form of learning assistance, tutorial classes, or transitional classes. Group 3 included students in general education junior secondary school classes. Group 4 included students who took honours classes in either English or Social Studies, in any of Grades 8, 9, or 10.

It is important to note that letter grades or percentage marks for Group 1 students were mostly unavailable. Once placed in a special education program, these students no longer received letter grades for academic courses, but were marked according to their progress on Individual Educational Plans (IEPs). This may have contributed to a possible decrease in the group variability of the percentage marks used in this study, posing a threat of reduced reliability.

Data Analyses

Descriptive statistics are presented for each analysis completed. Calculations were generated using SPSS computer applications (SPSS version 7.5 for Windows, 1996).

Because validity coefficients are directly related to the reliability of the measures used, a preliminary analysis of the reliability of junior secondary school marks was undertaken. The

reliability of district CBM measures was presented in the School District #57 guidebook (1996b). Appendix D provides the stability and equivalence reliability tables pertinent to this study.

Data collected were then analyzed following normal procedures for examining criterion-related predictive validity. Pearson's product-moment correlation coefficients were calculated between the identified predictor variables (CBM reading and written expression scores) and the criterion measures (junior secondary school marks in English and Social Studies for Grades 8, 9, and 10).

Although the reliability of the available measures is accepted as a possible limitation of the study, it is also appropriate to eliminate errors of measurement from the validity coefficient through a correction for attenuation (Sax, 1997). Since errors of measurement can attenuate, or lower, correlation coefficients, a correction for attenuation was considered for the criterion variable to ensure that judgment of the predictive validity of CBM scores was not affected by the possible unreliability of junior secondary school marks. A validity coefficient corrected for attenuation on both variables (CBM scores and junior secondary school marks) was also examined. These procedures resulted in correlations based on estimates of true scores, providing maximum predictive validity coefficients.

An examination of the strength of the relationships between combinations of predictor variables and the criterion measures was carried out through a multiple regression analysis. This determined how well the combined predictor variables of CBM reading and written expression scores predicted the grade criterion measures for junior secondary school humanities courses. The relative contribution of each CBM predictor variable to the relationship was identified. Additional regression analyses examined how well Grade 9 or 10 marks could be predicted from Grade 8 marks in the same subject. Standard error of estimate calculations provided information about the degree of confidence that could be placed in the interpretation of the results.

Discriminant analysis procedures were used to provide additional CBM validation evidence related to differentiation among student groups. The analysis focused on whether there was significant differentiation between groups based on CBM scores. No special problems were posed by unequal sample sizes in the groups. According to Tabachnick and Fidell (1996), when examining multivariate normality, robustness is expected with at least 20 cases in the smallest group if there are five or fewer predictors. Post hoc tests of the least significant difference (LSD) added a more detailed examination of group differentiation to the discriminant analysis using WRC and WSC separately.

CHAPTER FOUR: RESULTS

This chapter presents results for analyses conducted. First, the reliability of secondary school marks is examined to confirm the appropriateness of their use as criterion measures. Then the criterion-related validity of CBM is presented using Pearson's product-moment correlations. Corrections for attenuation demonstrate the maximum theoretical validity possible. Next, regression analyses provide a measure of the strength of predictive relationships found. The chapter concludes with additional validity evidence gathered from discriminant analysis and post hoc procedures to show the usefulness of CBM for differentiation between student performance groups.

Reliability of Criterion Measures

A preliminary analysis of the reliability of year-end junior secondary school marks was conducted. Table 1 shows the means, standard errors of the means (*SE*), and standard deviations (*SD*) for all marks awarded in English and Social Studies in Grades 8, 9, and 10. Data indicated general overall consistency in the average marks awarded between courses and among grades. Gradual increases in the standard errors of the means were noted as grade levels increased, and standard deviations were most pronounced at the Grade 10 level in both course areas.

Table 1

Descriptive Statistics for Secondary Marks

Course	*n*	Mean	*SE*	*SD*
Eng 8	430	71.3	0.68	12.5
Eng 9	422	69.1	0.75	13.6
Eng 10	212	70.1	0.99	14.4
SS 8	428	70.9	0.71	13.2
SS 9	415	68.4	0.75	13.8
SS 10	214	68.5	1.10	15.7

Note. Values shown calculated in percentages.

A correlational analysis provided stability and equivalence information on how well English and Social Studies marks related to each other for the 182 students who received marks in each course for all three grades. Although there are no set rules for what constitutes a minimally acceptable value for reliability coefficients, many standardized achievement test manuals report coefficients ranging in the .80s and .90s for reliability estimates of equivalence (Crocker and Algina, 1986). Secondary mark values shown (see Table 2) came close to this range, with a high degree of consistency for within-subject correlations (.75, .75, and .79 for English; .73, .74, and .74 for Social Studies).

These patterns were very regular, showing stability across grades in the same courses.

Table 2

Correlation Matrix for English and Social Studies Marks

	EN8	EN9	EN10	SS8	SS9	SS10
EN8	1.00	.75	.75	.82	.71	.72
EN9		1.00	.79	.69	.71	.71
EN10			1.00	.66	.72	.75
SS 8				1.00	.74	.73
SS 9					1.00	.74
SS 10						1.00

With the one exception of Grade 9 English and Social Studies, coefficient values for within-grade correlations (.82, .71, and .75) are higher than across grade and across subject values (.66, .69, .71, .71, .72, and .72). Overall, subject marks behaved as expected, with the least amount of correlation noted between marks from different courses in different grades.

As suggested by Sax (1997), all assessment tools should be internally consistent. Measures of internal consistency tend to be higher than long-term stability or equivalence coefficients, and are often used as their upper limit. Coefficient alpha was used to estimate the internal consistency of composite subject

scores. This statistic was calculated for English marks only, Social Studies marks only, and all marks combined. Resulting values for coefficient alpha indicated a high degree of same-course internal reliability (r = .90 for English, r = .90 for Social Studies), and an even higher degree of internal reliability (r = .94) for all marks considered together.

Using the Spearman-Brown formula for split-half reliability reported by Sax (1997), the correlation coefficient between subject marks was .95. This is another measure of the internal consistency of English marks and Social Studies marks combined, supporting the coefficient alpha calculation of .94. These results indicated a very predictable degree of similarity between subject marks, with all marks acting as measures of similar academic achievement traits.

Criterion-Related Validity

Choice of CBM Measures for Analyses

The CBM reading probe score used to measure reading fluency was Words Read Correctly (WRC). Total Words Written (TWW) and Words Spelled Correctly (WSC) were the two written expression fluency scores collected from writing probes. Initial correlation analyses were computed among the three identified CBM measures. Analyses for combined Grade 6 and Grade 7 year-end scores indicated coefficients of .43 for the WRC and WSC

correlation, and .38 for the WRC and TWW correlation. An almost perfect relationship was found (r = .99) between TWW and WSC. This confirmed results reported in the CBM norming guidebook (School District #57, 1996b) for these grades (r = .98 in Gr. 6, r = .99 in Gr. 7).

The decision was made to use WRC and only one of the written expression measures for further analyses, as variables that are too highly correlated can create collinearity problems in an analysis. Redundant variables tend to inflate the size of error terms, render unstable matrix calculations, and weaken the analysis (Tabachnick & Fidell, 1996). The WSC scores were chosen as they are generally considered more stringent measures of basic writing skill than TWW scores. The two measures, WRC and WSC, were chosen as the general indicators of achievement for Grade 6 and Grade 7.

Relationship Between CBM and Secondary School Marks

Means and standard deviations for student performance on year-end WRC and WSC probes were calculated. For combined Grade 6 and 7 WRC, the mean was 131 with a standard deviation of 42 (n = 639). For combined Grade 6 and 7 WSC, the mean was 58 with a standard deviation of 17 (n = 637). Means and standard deviations for junior secondary marks have already been presented in Table 1.

Table 3 shows Pearson product-moment correlations generated between CBM predictor variables and secondary school mark criterion measures. The resulting range of correlations indicated a definite, positive, and significant relationship between reading and written expression CBM scores and junior secondary school English and Social Studies marks, even over extended time periods.

Table 3

Correlations Between CBM Scores and Secondary School Marks

Secondary Course	WRC	WSC
EN 8	.46**	.34**
EN 9	.38**	.29**
EN 10	.32**	.28**
SS 8	.39**	.24**
SS 9	.36**	.16*
SS 10	.30**	.21*

Note. *Correlations significant at $p < .005$ level.
**Correlations significant at $p < .0005$ level.

In every case, WRC correlated more positively with course marks than did WSC. At each grade level, both the WRC and WSC measures correlated more positively with English than with Social Studies. With only one exception (SS9 and WSC),

correlations were highest in Grade 8 and gradually decreased over time, as would be expected. Although both CBM measures show predictive validity, the WRC measure was found to be the better general indicator of student achievement in junior secondary school Humanities courses, and more specifically in secondary English courses.

According to Cohen's (1992) effect size guidelines, all coefficient values represented non-trivial effect sizes beyond the .1 value considered small, but important. Effect sizes for correlations with WRC were medium and higher. Small to medium effect sizes were identified for correlations with WSC. The significance level of all correlation coefficients demonstrated a greater than 99.5% degree of confidence for rejection of the null hypothesis. These results indicated the importance of the validity coefficients, supporting the predictive validity of CBM scores as indicators of future academic achievement.

Corrections for Attenuation

Even though calculated validity coefficients were not considered trivial in their unattenuated state, corrections for attenuation were applied to examine the influence of errors of measurement (Sax, 1997). A correction for attenuation on the criterion reduced the effects of errors of measurement related to the reliability of secondary course marks. Both secondary course marks and CBM scores were corrected for attenuation to

provide the maximum predictive validity coefficient theoretically possible if all error was eliminated from both correlation variables.

Table 4

Validity Coefficients Corrected for Attenuation

CBM	Secondary Course	*r*	Corrected *r* (Crit)	Corrected *r* (Crit/Pred)
WRC	EN 8	.46	.53	.58
	EN 9	.38	.44	.48
	EN 10	.32	.37	.40
	SS 8	.39	.45	.49
	SS 9	.36	.42	.45
	SS 10	.30	.35	.38
WSC	EN 8	.34	.45	.50
	EN 9	.29	.39	.43
	EN 10	.28	.37	.41
	SS 8	.24	.32	.36
	SS 9	.16	.22	.24
	SS 10	.21	.28	.31

Table 4 shows corrected values using secondary course marks as the criterion measure and the CBM measures of WRC and WSC as predictor measures. In each case, the increase to coefficient values by correcting the criterion for attenuation was greater than the increase to the corrected criterion coefficient value when either of the CBM predictors were added to the equation.

Correcting both the secondary course marks and CBM scores for attenuation resulted in overall gains on validity coefficients of .08 to .12 for WRC and .08 to .16 for WSC. Both measures resulted in coefficients of .40 or greater for correlations with English course marks at all grade levels. Correlations with Social Studies course marks were similar for WRC, but remained lower for WSC.

Strength of Predictive Relationships

To further examine the strength of the relationships between CBM measures and secondary school marks, a series of regression analyses was conducted. Data were examined for residuals indicating outlier cases prior to regression analyses, because cases poorly fit by the regression equation often have an undue impact on analysis results. Using the criterion established by Tabachnick and Fidell (1996), seven outlying cases with standardized residuals in excess of ± 3.3 were removed from the secondary course marks analysis. All other cases were retained for regression calculations.

Regression results indicated how well the combined variables of CBM reading and written expression predicted secondary school mark criterion measures, and how large a contribution was made by the separate CBM variables to the overall predictive accuracy. For the multiple regression portion of this analysis, WRC and WSC were entered into the equation as predictor variables. Grade 8 English and then Grade 8 Social Studies were examined as separate criterion measures. Results are shown in Table 5.

Table 5

Contribution of CBM Variables to Overall Predictive Accuracy

Criterion	Predictor	Adjusted R^2	*SEE*	*F*
Eng 8	WRC	.21	12	112.71
	WSC	.11	13	53.83
	WRC + WSC	.24	12	65.01
SS 8	WRC	.15	13	76.79
	WSC	.05	14	22.74
	WRC + WSC	.16	13	39.65

Note. All *F* values significant at $p < .0005$ level.

The proportion of total variance in the junior secondary school scores that was accounted for by CBM performance is indicated by the adjusted R^2. The adjustment reduces the

influence of overestimations to R^2 due to chance fluctuations. Both WRC and WSC made statistically significant contributions to each regression equation, but when looking at the combined CBM measures, there was only a small additional proportion of variance accounted for by WSC (.03 for Eng 8 and .01 for SS 8). This again confirmed that WRC was the better predictor of marks for both Humanities courses.

Table 6

Predictive Accuracy of Secondary School Marks

Predictor	Criterion	Adjusted R^2	*SEE*	*F*
Eng 8	Eng 9	.59	9	570.79
	Eng 10	.56	9	249.00
SS 8	SS 9	.60	10	589.93
	SS 10	.58	10	269.89

Note. All *F* values significant at $p < .0005$ level.

The standard error of the estimate (*SEE*) provides a measure of the spread of errors, or the degree of fluctuation to be expected around true score values, that denotes the predictive ability of the variables examined. Results indicated that using either WRC alone or combined WRC and WSC measures, Grade 8 English course marks could be predicted within an interval of

the true score ± 12% and Grade 8 Social Studies course marks could be predicted within an interval of the true score ± 13%.

Additional regression analyses were employed to determine how well Humanities school marks in Grade 8 predicted marks awarded in Grades 9 and 10 for the same course areas. These results, shown in Table 6, were then compared to the previous analysis that used CBM measures as the predictors.

As expected, correlating marks in the same course area from year to year produces considerably higher adjusted R^2 values than correlations between different courses and different types of measures. However, in comparison to the predictive accuracy of the CBM measures from the previous analysis, the standard error of the estimate is not greatly improved. By predicting secondary school course marks from marks in the same course area at a lower grade level, an interval of the true score ± 9% and 10% is created, as compared to ± 12% to 14% range using one-time CBM scores from the end of Grades 6 and 7.

Additional Validity Evidence

Discriminant Analysis of Group Membership

This analysis examined the usefulness of a combination of CBM scores for differentiating between student groups classified into four categories of academic achievement based on their

secondary school program placements. Once all cases with missing values were deleted, the data set included 454 cases. No cases were deleted as outliers, as no values extended beyond the recommended ±3.3 range for standardized residuals (Tabachnick & Fidell, 1996). Table 7 displays group statistics.

Table 7

Discriminant Analysis Group Statistics

Group	*n*	WRC Mean	WRC *SD*	WSC Mean	WSC *SD*
1	21	84	33	39	13
2	42	102	35	51	18
3	328	136	39	59	15
4	63	158	33	67	16
Total Group	454	133	41	58	16

Group 1 included students placed in some form of alternate special education class, Group 2 included students receiving other remedial support, Group 3 included students in general education classes, and Group 4 included students in honours classes.

A normal distribution is indicated when 68% of the sample falls with the ± 1*SD* range and 16% of the remainder fall in each

of the upper and lower ranges. The groupings in this study fit closely with this model, identifying 72% of the sample in the average, or general education range, 14% (5% in Group 1 and 9% in Group 2) in the lower remedial support range, and 14% in the upper honours placement range.

Residuals (errors of prediction) were examined on a normal probability plot where expected normal values were plotted against actual normal values. The residuals plot and group sizes met requirements for assuming multivariate normality. This ensured that analysis results would not be distorted.

Tests of equality of group means for each CBM measure in the discriminant analysis indicated significant ($p < .0005$) overall results, with $F(3, 450) = 31.78$ for WRC, and $F(3, 450) = 19.91$ for WSC. These results demonstrated that there was reliable separation of the four program placement groups using both CBM measures.

With more than two groups, there can be more than one way to combine the predictors to differentiate between the groups (Tabachnick & Fidell, 1996). In this discriminant analysis summary, two predictor combinations (functions) for separating the groups from each other were extracted. The eigenvalue of the first function identified it as a dominant factor, accounting for 98% of the between-group variance contributed to the discriminant function equation. A canonical correlation of .45

for the first function indicated the proportion of variance shared between predictors and groups in that function, as compared to the canonical correlation of .08 for the second function. These results show that the first function provided the most discriminating power, with little information added by the second function.

With both functions included in the discriminant analysis equation, a highly reliable degree of relationship between groups and predictors was indicated by $\chi^2(6) = 105.1$, $p < .0005$. A reliable relationship was not supported by removal of the first function, as indicated by $\chi^2(2) = 2.5$, $p = .28$. These results confirmed that the combined CBM variables could be used to reliably predict group membership, with the first function providing the best separation among groups.

Group Differentiation by Separate CBM Measures

Post hoc tests of the least significant difference (LSD) were run to determine which particular group means differed significantly from the others. This provided a more detailed examination of group differentiation using WRC and WSC scores, as displayed in Table 8. In almost all group comparisons on both measures, the mean difference between groups was significant at the $p < .005$ level. The one exception was the differentiation between Groups 1 and 2 on the WRC measure ($p = .076$). This one value, however, based on a sample of only 21 students, almost

met the $p < .05$ significance. Both Groups 1 and 2 remedial program students were effectively separated from students in general education classes, as were the Group 4 honours students.

Table 8

Post Hoc (LSD) Tests of Group Comparisons

Predictor Variable	Group (I)	Group (J)	Mean Diff. (I-J)	Std. Error	Sig. (p)
WRC	1	2	-18	9.97	.076
		3	-51	8.39	.000
		4	-74	9.38	.000
	2	3	-34	6.11	.000
		4	-55	7.40	.000
	3	4	-22	5.09	.000
WSC	1	2	-12	4.08	.004
		3	-19	3.43	.000
		4	-27	3.84	.000
	2	3	-7	2.50	.003
		4	-15	3.04	.000
	3	4	-8	2.10	.000

In order to look at the differentiation of all students receiving remedial support from general education or honours students, Groups 1 and 2 were combined and the post hoc (LSD) tests were repeated. Using these results, all comparisons between groups with both WRC and WSC scores were significant at the $p < .0005$ level. The ability of CBM to separate performance groups

was supported at a higher level of confidence with a three-group model than with a four-group model. A comparison of the three-group model to the four-group model in a repeat of the discriminant analysis indicated slight differences in statistical values, but all still at the $p < .0005$ level.

Overall group differentiation ability was supported with both models of grouping. The three-group model displayed more significant differences between groups. The four-group model provided increased sensitivity with a loss of significance on one comparison.

CHAPTER FIVE: DISCUSSION

Summary

Extensive previous research has already found CBM to be a valid assessment tool in all steps of a problem-solving model of service delivery. In many cases, criterion-related studies compared CBM to traditional commercial standardized measures, reporting positive relationships that demonstrated the utility of CBM as a reliable and valid indicator of student performance in basic skills. Its effectiveness in determining that a student's skills are sufficiently different from those of other students has also validated the use of CBM for screening and eligibility decisions.

This study was conducted to validate the use of locally developed CBM norms as a basis for resource allocation decisions in School District #57. Secondary school marks were selected as widely available school-based criterion measures. Evidence from this investigation confirms the hypothesis that there is a predictive relationship between reading and written expression CBM scores and future teacher-assigned secondary marks in Humanities courses, indicating that CBM is a valid measure of student achievement. The hypothesis that CBM scores can identify a discrepancy between student groups with different levels of academic performance was also supported, verifying the

appropriateness of using CBM as one measure of student discrepancy from the norm.

Conclusions

The criterion-related validity of CBM scores was examined by confirming its ability to predict future academic achievement as measured by junior secondary school course marks. The first step verified the reliability of marks assigned, in order to validate their use as criterion measures.

Reliability results show very consistent patterns within same course areas across Grades 8, 9, and 10. Marks in the two separate courses within the same grades show more variability, although higher Grade 8 coefficient values likely reflect the combination of English and Social Studies in some schools. It can be seen that there is greater reliability within same course areas and within same grades than there is across different courses and different grades. These are the results that would be expected, showing that the secondary marks are behaving in a predictable manner contrary to what may be suggested in literature related to secondary school marks reliability.

Considering the many possible factors that influence the magnitude of reliability coefficients for secondary marks, lower values than the generally desired .80s and .90s might be indicated. In fact, actual stability coefficients obtained for

same course area and same grade correlations are not much lower, averaging in the mid .70s. In addition, internal consistency coefficients of .90 to .95 indicate a high degree of homogeneity for secondary marks, suggesting that they do indeed measure something very similar. These combined results indicate that even taking all possible error variance into account, secondary marks are more reliable than expected and can be used as reasonable criterion measures of general student achievement outcomes.

Validity predictions tend to be less accurate as the length of time between predictor and criterion measurements increases, and predictive validity coefficients tend to be lower than concurrent validity coefficients. As with reliability estimates, the possibility of chance intervening factors can lower the correlational relationships found. In this study, there were gaps of one, two and three years between the time CBM scores were collected and the collection of end-of year junior secondary school marks. In addition, CBM scores were obtained from one-time, short duration probes (one minute for WRC, three minutes for WSC).

In light of these factors, correlation coefficients with important effect sizes and high significance levels fit well with guidelines for demonstrating predictive utility. No conclusion was reached as to the one inconsistent coefficient

discovered for Grade 9 Social Studies and the WSC measure, registering a lower correlational value than Grade 10 Social Studies. In spite of this discrepancy, correlation results corroborate validity claims in the literature and refute the belief of some teachers that CBM measures are unlikely to reflect student performance.

Coefficient values corrected for attenuation reduce the effect of measurement errors, giving a closer estimate of the maximum predictive validity theoretically possible. Although not necessary to prove the utility of CBM, corrected values for both measures strengthen the evidence of their ability to predict future achievement. This further supports their use as indicators of student performance in the basic skills of reading and written expression.

The adequacy of CBM scores and junior secondary school scores as measures of the same construct of academic achievement was examined. Standard errors of estimates from regression analyses show the comparative predictive power of CBM scores. These brief, simple measurements taken one, two or even three years before the awarding of the secondary course marks were almost as effective as indicators of later Humanities course marks as were marks from those same course areas calculated from an entire term's work at a lower grade level. It is possible that these results may have been even more substantial if the

CBM data had been collected from more than one probe measurement, as is recommended when using CBM for assessment purposes.

Correlation results, combined with evidence of the predictive power of CBM, show that these measures can contribute useful information when making decisions about students based on their academic performance. Additional support for the use of CBM when making decisions comes from the discriminant analysis results. Students who ultimately received remedial support in secondary school were differentiated from general education students on the basis of CBM scores collected in Grade 6 and Grade 7. Students entered into honours programs were also reliably separated from other student groups.

The differentiation between student performance groups based on CBM scores means that School District #57 personnel can be confident that the use of CBM as one measure to screen students for the allocation of additional resources is appropriate. Results corroborate previous research describing the discriminative validity of CBM when comparing low-achieving, learning-disabled, or mildly-handicapped groups to general education groups (Schendel & Binder-Reschly, 1989; Shinn & Marston, 1985). The ability of the measure to differentiate students in the low-achieving range of the School District #57

distribution supports the desired assessment consequence of interventions and resources being applied where needed.

In the discriminant analysis function, group membership by a combination of the two CBM scores is determined, with reliable separation of the four student performance groups. The magnitude of the differences between both low-achieving groups and the general education group was also significant in post hoc tests of the separate CBM measures. The mean difference between the two levels of remedial grouping for low achieving students was less significant, however, particularly on the WRC measure. This indicates that similarities in reading performance are shown by some of the students in both low-achieving groups.

The small sample of students in the special education placement group (Group 1) may have affected these post hoc results. Another possible contributing factor is the unknown amount of error generated in this small group by only one reading probe administration. In practice, school district personnel are directed to acquire multiple CBM samplings and use the median score when assessing performance levels in relation to the CBM norms. Cut-off scores in the norms then allow for the determination of severe performance discrepancies for individuals. The fact that the CBM scores collected were sufficient to differentiate between student performance groups

does not suggest that overgeneralizing about individual student performance on the basis of only one measurement is acceptable.

Limitations

Cautions for the interpretation of longitudinal data have been acknowledged, as has the concern about the possible subjectivity or lack of consistency related to secondary course marks. In spite of these qualifications, secondary course marks were selected as criterion measures that are important to teachers, students, and parents. It is possible that more accurate school-based measures of student performance could be identified, but it is unlikely that they would be as widely available for the large student sample followed in this study.

Many factors potentially impact on reliability and validity results as sources of error. These include the length of time between marks awarded, differences between school programs and teachers awarding marks, reduced group variability through natural attrition or artificial restrictions of range, and changes within students themselves. Decisions about what data to include in the study and the organizational methods used may have reduced the sensitivity of analysis results. The secondary course marks included, the selection of specific CBM scores from the norming project, and the specific analysis questions examined are all examples of the research choices made.

The acceptance and application of study results will be determined by their interpretation. Moderate validity coefficients for WRC and low to moderate coefficients for WSC, all statistically significant, show that CBM measures can contribute to the decisions made about students. Confidence in their predictive ability as indicators of student performance can be increased with the understanding that predictive validity coefficients rarely exceed the moderate range.

The results obtained provide generalized group information. They do not address the concerns of teachers related to specific outliers. Outliers are the students with unexpectedly high or low CBM scores that do not accurately reflect their overall reading or written expression ability. In such instances where there is considerable discrepancy between CBM performance and teacher judgment of student performance, further assessment with alternative measures is recommended. Even if a measure is considered an important indicator, it does not mean that it should be the only source of information collected. District policy reflects this belief in assessment guidelines for resource eligibility guidelines.

Implications for Future Research

The lack of acceptance of the CBM oral reading fluency measure is a face validity issue (Shinn et al, 1992). The

written expression measures used to create local norms face a similar problem. Face validity is not always improved by reports of technical adequacy. In personal communication on this subject, S. L. Deno (December 2,1999) suggests the collection of supplementary information along with CBM scores. Teachers feel more comfortable about results with the addition of test tasks that "look" more valid to them, and with experience begin to focus more on the CBM scores. A recommendation for further research to examine this controversial issue is related to the work of Hintze et al. (1997), who reported on studies that found oral reading fluency to correlate more strongly with teacher judgment than any commercially available norm-referenced reading measures examined. A district study comparing CBM scores on reading or writing measures with teacher evaluations of reading or writing competence would give teachers a personally relevant measure of CBM validity.

Research related to CBM written expression measures is recommended. Although the TWW and WSC measures are sufficient for the development of norms to detect significant differences in student performance, they are not complete measures of overall writing competence, and do not demonstrate the same level of validity shown by the WRC measure. The additional measure of Correct Word Sequences (Deno, Marston, & Mirkin, 1982), not widely used in the district, may provide more useful

diagnostic information for teachers. Alternatively, other means of scoring written expression probes (currently being researched locally) are worth considering.

Implications for Practice

With the validation of CBM presented in this study, the implication is that both teacher and district confidence in CBM for its current application is justified. As an indicator of future academic achievement, CBM results can help to identify problems that warrant further investigation. A consequence of using CBM for screening and eligibility purposes is increased effectiveness in the allocation of time and resources. Appropriate intervention plans directly related to the curriculum can be considered earlier. Initial assessments can then be extended into progress monitoring.

In a discussion of their own research results, Schendel and Binder-Reschly (1989) caution against the potential danger of limiting the application of CBM solely to screening and eligibility decision-making because evidence collected supports that purpose. They argue that the true value of CBM, its sensitivity to small changes in instruction over short periods of time, should not be overlooked. Maximized student learning through the planning, implementation and monitoring of interventions specific to the curriculum is facilitated through

the use of CBM. Expanding the use of CBM to include these originally intended purposes in School District #57 is a practical goal worth considering.

In spite of previous and current research confirming the technical adequacy of CBM, many teachers continue to doubt its ability to measure reading and writing competence. The use of CBM has not been completely accepted by teachers due to misuse of procedures outlined in the guidebook, misunderstanding of the intent of the norms developed, and confusion related to errors in the interpretation of CBM results. Additionally, alternate purposes and procedures for CBM have not been widely explained. In particular, the use of CBM for monitoring progress and evaluating program effectiveness needs to be more fully explored. Further training and discussion of CBM procedures for other applications within the problem-solving model would be helpful, with information presented to help develop an understanding of what these measurement tools can and cannot do.

Choate et al. (1992) discuss the criteria of a good test. Foremost, a good test is one that best meets the purpose of the assessment, and should not be used if it does not measure what it is supposed to measure. They outline five desirable test characteristics:

1. Acceptable technical characteristics (validity and reliability)

2. Easy to administer

3. Easy to score

4. Not too time consuming

5. Pleasant for the student

They add that assessment methods used to make decisions about instruction must also be capable of detecting small performance changes and be amenable to frequent administration. An additional practical consideration is a reasonable cost related to materials and personnel involved. CBM is an assessment procedure that meets these criteria.

This study provides local validation for the use of CBM in School District #57. Along with its other desirable qualities, the technical adequacy reported makes it an attractive element in the assessment repertoire of school personnel. Evidence of the value of CBM as an indicator of academic achievement and its ability to differentiate between student performance groups adds to the technical data reported in the guidebook for CBM use. It also contributes new information to the substantial body of research supporting the utility of CBM procedures.

REFERENCES

Allinder, R. M. (1996). When some is not better than none: Effects of differential implementation of curriculum-based measurement. *Exceptional Children, 62,* 525-535.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

British Columbia Ministry of Education, Educational Programs (1990). *The Intermediate Program: Learning in British Columbia* (Response Draft). Victoria, BC: Province of British Columbia.

British Columbia Ministry of Education. (1994). *Guidelines for student reporting for the Kindergarten to Grade 12 education plan.* Victoria, BC: Ministry of Education.

Choate, J., Enright, B., Miller, L., Poteet, J., & Rakes, T. (1992). *Curriculum-based assessment and programming.* Needham Heights, MA: Allyn and Bacon.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155-159.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L. (1989). Curriculum-based measurement and special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1-17). New York: Guilford Press.

Deno, S. L. (1992). The nature and development of curriculum-based measurement. *Preventing School Failure, 36*(2), 5-42.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children, 48,* 368-371.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Elliot, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review, 26*(2), 224-233.

Espin, C. A., & Deno, S. L. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education, 14*, 47-59.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62,* 497-514.

Fuchs, L. S. (1989). Evaluating solutions: Monitoring progress and revising intervention plans. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 155-183). New York: Guilford Press.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232-242.

Fuchs, L. S., Deno, S. L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision-making. *Diagnostique, 8,* 135-149.

Fuchs, L. S., & Fuchs, D. (1986). Curriculum-based assessment of progress toward long-term and short-term goals. *Journal of Special Education, 20,* 69-82.

Fuchs, L. S., & Fuchs, D. (1990). Traditional academic assessment: An overview. In R. A. Gable & J. M. Hendrickson (Eds.), *Assessing students with special needs* (pp. 1-14). New York: Longman.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*, 20-28.

Gersten, R., Keating, T., & Irvin, L. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children, 61,* 510-519.

Good, R. H. III, & Jefferson, G. (1998). Contemporary

perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61-88). New York: Guilford Press.

Guskey, T. R. (1994). Making the grade: What benefits students? *Educational Leadership, 52*(2), 14-20.

Hedekar, L. (1997). *The effects of month of birth and gender on elementary reading and writing fluency scores using curriculum-based measurement.* Unpublished master's thesis, University of Northern British Columbia, Prince George, British Columbia, Canada.

Hintze, J., Shapiro, E., Conte, K., & Basile, I. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review, 26,* 535-553.

Holmes, M. (1993). *The educator's guide to student evaluation.* Toronto, ON: The Ontario Institute for Studies in Education.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp.18-78). New York: Guilford Press.

Marston, D. B., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52,* 266-276.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18,* 109-117.

McDaniel, E. (1994). *Understanding educational measurement.* Madison, WI: WCB Brown & Benchmark.

Messick, S. (1989). Validity. In R. L. Linn (Ed), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Principles for fair student assessment practices for education in Canada. (1993). Edmonton, AB: Joint Advisory Committee. (Available from the Joint Advisory Committee, Centre for Research in Applied Measurement and Evaluation, 3-104 Education Building North, University of Alberta, Edmonton, AB, T6G 2G5).

Reschly, D., & Grimes, J. (1990). Best practices in intellectual assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 425-439). Washington, DC: National Association of School Psychologists.

Salvia, J., & Ysseldyke, J. (1991). *Assessment* (5th ed.). Boston: Houghton Mifflin.

Sax, G. (1997). *Principles of educational and psychological measurement and evaluation* (4th ed.). Belmont, CA: Wadsworth Publishing.

Schendel, J., & Binder-Reschly, M. (1989). *Criterion validity of curriculum-based assessment and correlation with teacher ratings and ITBS scores.* Des Moines: Iowa Department of Education, Bureau of Special Education. (ERIC Document Reproduction Service No. ED 318 152)

School District #57. (1994). *Report of the school services task force to the board of school trustees: School District #57 (Prince George).* Prince George, BC: School District #57.

School District #57. (1995). *Curriculum based measurement norming project manual.* Prince George, BC: School District #57.

School District #57. (1996a). *School support services manual: Practices, organization, principles.* Prince George, BC: School District #57.

School District #57. (1996b). *Guidebook for the use of curriculum based measurement in School District #57.* Prince George, BC: School District #57.

Shapiro, E. S. (1989). *Academic skills problems: Direct assessment and intervention.* New York: The Guilford Press.

Shepard, L. A. (1993). Evaluating test validity. In L. Darling-

Hammond (Ed.), *Review of Research in Education* (Vol. 19, pp. 405-450). Washington, DC: American Educational Research Association.

Shinn, M. R. (1989). Identifying and defining academic problems: CBM screening and eligibility procedures. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp.1-17). New York: Guilford Press.

Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford Press.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big Ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 1-31). New York: Guilford Press.

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21,* 459-479.

Shinn, M. R., & Habedank, L. (1992). Curriculum-based measurement in special education problem identification and certification decisions. *Preventing School Failure, 36*(2), 11-15.

Shinn, M. R., & Hubbard, D. D. (1992). Curriculum-based measurement and problem-solving assessment; Basic procedures and outcomes. *Focus on Exceptional Children, 24*(5), 1-20.

Shinn, M. R., & Marston, D. (1985). Differentiating mildly handicapped, low-achieving and regular education students: A curriculum-based approach. *Remedial and Special Education, 6,* 31-45.

Shinn, M., Nolet, V., & Knutson, N. (1990). Best practices in curriculum-based measurement. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 287-307). Washington, DC: National Association of School Psychologists.

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of

classroom performance. *Journal of Learning Disabilities, 19,* 545-552.

Tabachnick, B., & Fidell, L. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College Publishers.

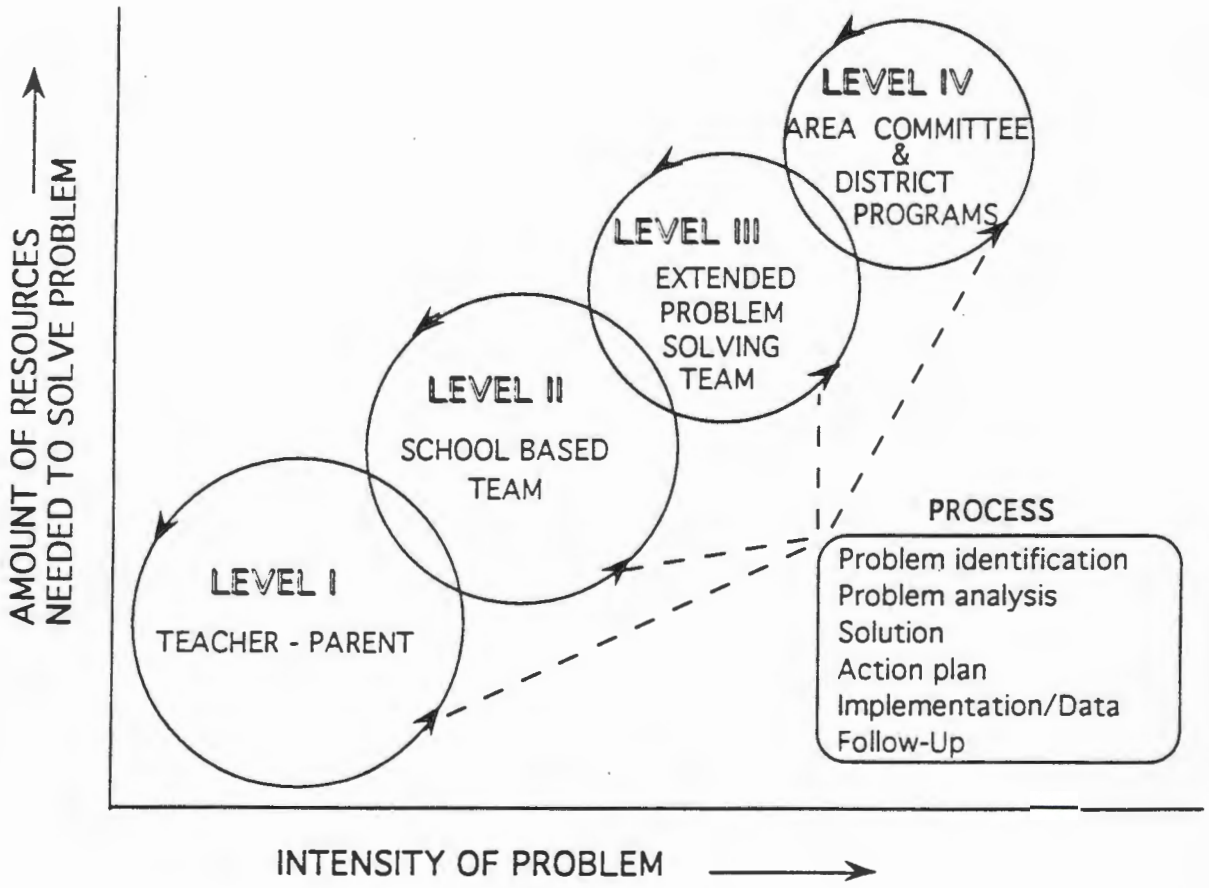
Thorndike, R. M (1997). *Measurement and evaluation in psychology and education* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Tilly, W. D. III, & Grimes, J. (1998). Curriculum-based measurement: One vehicle for systematic educational reform. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 32-60). New York: Guilford Press.

APPENDIX A

School District #57 Problem-Solving Model

PROBLEM SOLVING APPROACH



* Adapted from Heartland Education Agency 11
Division of Special Education - Iowa

APPENDIX B

Reported Reliability from School District #57 CBM Project

Equivalence and Stability Correlation Coefficients

Pearson Correlation between *Total Words Written* and *Words Spelled Correctly*			
Grade	OCTOBER	JANUARY	APRIL
1	-	-	.92
2	.92	.91	.95
3	.96	.96	.96
4	.96	.97	.98
5	.96	.97	.96
6	.97	.98	.98
7	.99	.99	.99

Pearson Correlation between *Total Words Written* and *Words Read Correctly*			
Grade	OCTOBER	JANUARY	APRIL
1	-	-	.45
2	.31	.40	.48
3	.40	.42	.38
4	.38	.38	.38
5	.32	.32	.34
6	.32	.32	.38
7	.39	.39	.37

Pearson Correlation for *Total Words Written* Scores between Norming Periods			
Grade	$r_{Oct-Jan}$	$r_{Jan-Apr}$	$r_{Oct-Apr}$
1	-	-	-
2	.55	.68	.48
3	.63	.67	.53
4	.67	.69	.61
5	.59	.62	.57
6	.64	.68	.62
7	.60	.64	.61

Pearson Correlation for *Words Read Correctly* Scores between Norming Periods			
Grade	$r_{Oct-Jan}$	$r_{Jan-Apr}$	$r_{Oct-Apr}$
1	-	-	-
2	.85	.84	.81
3	.89	.89	.86
4	.81	.80	.77
5	.85	.84	.83
6	.88	.85	.81
7	.86	.86	.86

*Adapted from Guidebook for The Use of Curriculum Based Measurement in School District #57 - Draft: November, 1996

APPENDIX C

Preliminary Thesis Outline

Submitted for School District Approval

Curriculum Based Measurement as an Indicator of Academic Achievement

Preliminary Thesis Outline

Saima Fewster

Background:

In the 1995-1996 school year, School District #57 developed local norms for Curriculum Based Measurement (CBM) scores in reading and writing. These norms were compiled in a guidebook for use by teachers in the problem-solving process to develop effective interventions for students. The norms allow for performance comparisons between students and same-grade peers within the district. As a learning assistance teacher, I frequently use CBM as one assessment tool for measuring student performance in reading and writing. It provides information about the severity of the problem for students with learning difficulties and evidence of a severe discrepancy when they are compared to their peers. This is one diagnostic criterion that is accepted when identifying a student for district-level support.

Research Purpose:

Although the district has endorsed and encouraged the use of CBM as an assessment tool, many teachers remain unconvinced that what is being measured can give an accurate indication of a student's level of reading and writing performance. In spite of the reliability information presented in the norming manual and reports of validity grounded in prior research, professionals have difficulty trusting the data provided by these seemingly simple little measures. There is currently no validity research related to the local data and I believe that the availability of this information would lend support to the use of CBM by teachers.

The intended purpose of my thesis research will be to examine the validity of CBM as a measure of student performance. In order to do this, I want to determine the nature of the relationship between CBM scores collected during the norming project and a selected criterion measure, specifically letter grades received by those students in subjects with high reading and writing components. I expect that my results will show that, in general, students with the lower CBM scores will have lower letter grades and students with the higher CBM scores will have higher letter grades. I want to see if further evidence can be provided for teachers to show the value of CBM as an assessment tool in identifying students in need of academic support. Further, I hope that confirming the validity of CBM will prompt its use by more teachers as a worthwhile method for the ongoing evaluation of intervention strategies in place for students.

Method:

My validity study will be based on approximately 600 students who were in Grades 6 and 7 at the time of the CBM project. Those students will have received up to three years of secondary letter grades by the end of this school year, providing me with longitudinal data that will make it possible to calculate predictive validity for the CBM scores using correlational

analyses. Several multi-variate analyses are possible, such as multiple regression, to determine the relationship between a number of presenting variables. As I become more involved in my study, the value of further analyses to meet my purpose will be explored. My thesis supervisor at UNBC is Dr. Peter MacMillan, who was involved in processing the data from the CBM project and developing the technical information for the guidebook. He has tentatively approved this topic for my thesis.

In order to carry out this study, I now need the approval of the school district, and its assistance in gathering student marks. This will involve allowing me the use of student scores from the original CBM project, currently in the possession of Peter MacMillan. It will also require locating the students who were in Grades 6 and 7 at the time of the project and gaining access to high school files at the end of this school year to examine their grade records. Letter grades could be collected by computer printout if possible, or through my manual searching of PR files as necessary.

Ethical Considerations:

Procedures to safeguard and maintain confidentiality of the data will be in place. Student records will be searched only by me or by designated school personnel. There will be no direct contact with students, and normal testing or grading procedures by the school will not be affected by the study. Student names will be coded and will not appear in any research findings. School notification and consent will be carried out as directed by the school district, explaining research procedures, purposes and provisions to protect the anonymity of selected students. If necessary, parent notification will also be requested prior to any data collection. Research results will be shared with university personnel and the school district.

Once the school district grants permission for me to carry out this study, I must submit my proposal to the UNBC Ethics committee for institutional approval of my research plan. An accompanying letter of consent from the appropriate school district administrator is required for this plan to be considered. UNBC procedures regarding ethics and confidentiality will be followed at all times during the study.

Summary:

This research plan has been generated by current practices in the school district. The results should be of interest to any school personnel using the CBM guidebook and procedures.

APPENDIX D

Letter to Secondary School Principals

May 01, 1999

(Principal's Name)
(Secondary School Name)

Dear

I am currently working as a Learning Assistance teacher at College Heights Elementary School, and am also working through the Education graduate program at UNBC for my Master's degree in Curriculum and Instruction. This letter outlines my current research project and is a request for your assistance in its completion. The project has been approved by Norm Monroe, Director of School Services for School District #57, who will be advised of any and all particulars of the project throughout its duration. The results of this study are of general interest to various school district personnel.

Background

In the 1995-1996 school year, all elementary schools in School District #57, in conjunction with UNBC participated in a project to develop local norms for Curriculum-Based Measurement (CBM) procedures. These norms are used to assist school staff in making academic decisions about student progress and help to determine the need or eligibility for resources beyond the school level. CBM has become a valuable tool for developing, monitoring and evaluating interventions within the problem-solving model adopted by the district.

Research literature reports extensive investigations of the technical adequacy of CBM as an assessment procedure, offering evidence to support its increasingly widespread use. While the reliability of our local norms has been well established and explained in the CBM guidebook, determining the validity of these norms will lend further credibility to their current use.

Current Study

In order to fulfill my Master's degree thesis requirements, I have proposed to carry out a criterion-related validity study of the School District's CBM norms. Because the CBM scores collected are specific to reading and writing, it is expected that subsequent student performance in the Humanities subjects will reflect the application of these skills. As we routinely use letter grades to report academic achievement, one measure of the validity of CBM scores will be to identify a positive correlation between the CBM scores and letter grades received by the students in their Humanities subjects since the norming year.

Method

To carry out this study, CBM data for students who were in Grades 6 and 7 at the time of the norming project will be provided from stored UNBC files. The identity and elementary school of each student will also be provided to facilitate the location of current high school placements. Letter grade information for these students will then need be to collected from high school records after the last reporting period in June, 1999. It is hoped that cumulative grade records for the students are available in permanent student files and are easily accessible through computer printouts, however I am willing to gather information manually if this is the most

reasonable approach. Data collection will be scheduled at the convenience of school personnel to minimize interference with normal end-of-year office procedures.

Ethics

This project will follow all UNBC research procedures and guidelines to safeguard and maintain information confidentiality. Student names will be coded once data is collected and will then be removed from all research documentation for further phases of the study. As data collection only involves examining existing school records, there will be no direct contact with the students and they will not be personally affected or identified by the study in any way. Any necessary parental consent or notification will be carried out according to school district and UNBC policy. The research proposal will be presented to the UNBC Ethics Committee for approval, and Dr. Peter MacMillan from the UNBC Education Department will supervise the project. Research results will be shared with university and school district personnel.

Summary

The success of this validation project will depend on your assistance in providing access to letter grade records for the students selected. Please advise Norm Monroe if you foresee any problems this study may involve for your school, or contact me if you have any questions, suggestions or concerns you would like to discuss. Assuming that all will go as planned, I will contact you later in May to make further arrangements.

This thesis plan has been generated by current practices in the school district. The results will be valuable to any school personnel using CBM procedures and the local norms guidebook. In addition to being able to fulfill my degree requirements, I am looking forward to the information that will be provided by this study, as it is relevant to my normal teaching and assessment practices.

Thank you in advance for your support and assistance.

Saima Fewster
College Heights Elementary School
964-4408
email: saima_fewster@fc.schdist57.bc.ca

Norm Monroe
Director of School Services
561-6800 Ext. 311
email: norm_monroe@fc.schdist57.bc.ca